

## Physicochemical bases for protein folding, dynamics, and protein-ligand binding

LI HuiMin<sup>1,3†</sup>, XIE YueHui<sup>2,3†</sup>, LIU CiQuan<sup>3,4</sup> & LIU ShuQun<sup>3,4,5\*</sup>

<sup>1</sup>*School of Mathematics and Computer Science, Yunnan University of Nationalities, Kunming 650500, China;*

<sup>2</sup>*Teaching and Research Section of Computer, Department of Basic Medical, Kunming Medical University, Kunming 650031, China;*

<sup>3</sup>*Laboratory for Conservation and Utilization of Bio-Resources & Key Laboratory for Microbial Resources of Ministry of Education, Yunnan University, Kunming 650091, China;*

<sup>4</sup>*Southwest Biological Diversity Laboratory, Kunming Branch of Chinese Academy of Sciences, Kunming 650223, China;*

<sup>5</sup>*Key Laboratory for Animal Genetic Diversity and Evolution of High Education in Yunnan Province, Kunming 650091, China*

Received May 7, 2013; accepted September 27, 2013; published online February 18, 2014

Proteins are essential parts of living organisms and participate in virtually every process within cells. As the genomic sequences for increasing number of organisms are completed, research into how proteins can perform such a variety of functions has become much more intensive because the value of the genomic sequences relies on the accuracy of understanding the encoded gene products. Although the static three-dimensional structures of many proteins are known, the functions of proteins are ultimately governed by their dynamic characteristics, including the folding process, conformational fluctuations, molecular motions, and protein-ligand interactions. In this review, the physicochemical principles underlying these dynamic processes are discussed in depth based on the free energy landscape (FEL) theory. Questions of why and how proteins fold into their native conformational states, why proteins are inherently dynamic, and how their dynamic personalities govern protein functions are answered. This paper will contribute to the understanding of structure-function relationship of proteins in the post-genome era of life science research.

**free energy landscape, entropy-enthalpy non-complementarity, ruggedness, driving force, thermodynamics, kinetics**

**Citation:** Li HM, Xie YH, Liu CQ, Liu SQ. Physicochemical bases for protein folding, dynamics, and protein-ligand binding. *Sci China Life Sci*, 2014, 57: 287–302, doi: 10.1007/s11427-014-4617-2

Proteins are important bio-macromolecules that are composed of one or more chains of amino acids. They perform a vast variety of functions *in vivo* and participate in virtually every process within a cell, including catalytic reactions in metabolism, response to stimuli during immune response, replicating DNA during cell division, and other crucial roles in signal transduction, molecular transport, cell adhesion, cell cycle, gene-expression regulation, and structural or mechanical functions in muscle and cytoskeleton. Therefore, understanding the functions of proteins is vital in the field

of life sciences.

The central dogma of protein structural biology is that the amino acid sequence contains all the information necessary for a protein to fold into its three-dimensional structure under the proper physiological/experimental environment, and that the structure is essential for protein function [1–3]. The high-resolution X-ray crystallography has brought about a revolution in the field of structural biology through producing numerous static structures, resulting in a surge of studies into the structure-function relationship of proteins [4]. However, because proteins are dynamic entities, the static unique structure or “folded state” of a protein ob-

†Contributed equally to this work

\*Corresponding author (email: shuqunliu@gmail.com)

tained from such experimental techniques cannot provide the final answer to its function. Indeed, the biological functions of proteins are rooted in their physical motions in the cell [5,6]. The physical motions consist of (i) protein folding process that results in the natively folded state; (ii) protein dynamics and molecular motions that occur within the folded conformational state; and (iii) the more specific function-related dynamics that occur during protein-ligand recognition and interaction.

Many methods have been developed to investigate protein dynamics on a variety of timescales at different resolutions [4]. Experimental approaches such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, Laue X-ray diffraction, cryo-electron microscopy, and small-angle X-ray scattering provide atomic-resolution snapshots or structural ensembles, which can be used to infer and/or obtain information about dynamics on slow timescales. Other classical biophysical techniques such as circular dichroism (CD), fluorescence resonance energy transfer (FRET), infrared spectroscopy, Raman spectroscopy, and electron paramagnetic resonance, despite being low-resolution and local-site methods, provide kinetic (i.e., the conformational transition) information over a large range of timescales. Computational approaches such as molecular dynamics (MD) simulation [7,8] and its derivative methods, including simulated annealing and replica exchange Monte Carlo techniques [9], force-field simplification models such as Gaussian network models [10] and Gō models [11], and enhanced sampling techniques such as metadynamics [12–14], umbrella sampling [15–18] and targeted and steered simulations [19–22], can provide not only complete information about protein dynamics at the atomic level (i.e., the precise positions of individual particle motions as a function of time), but also the underlying forces and corresponding energies that dictate such motions. Of particular note is that the enhanced sampling methods of MD can be used to calculate free energy differences and then reconstruct the FEL of a protein-solvent system. Computational simulations, coupled with improved computer power and advanced algorithms [23–25], have greatly facilitated protein dynamics studies, including studies on the protein folding mechanism, conformational space sampling, molecular motions, and conformational transitions, which in turn enhance significantly our understanding of the relationship between protein dynamics and functions.

The experimental and computational methods used for investigating protein folding and dynamics, as well as specific study cases about dynamics-function relationship of individual proteins, have been reviewed extensively elsewhere [4,6,26–29], and therefore are not covered in this review. Here we focus mainly on the fundamental physicochemical principles and/or mechanisms underlying the physical motions of proteins in an attempt to address the following questions: (i) Why can proteins fold rapidly into their native conformational states? What are the forces that

drive the folding process? How do they do so? (ii) Why are proteins inherently dynamic? What are the factors that dictate the protein dynamics? (iii) What are the driving forces for protein-ligand binding? How do they drive the recognition and binding processes? To address these questions the concepts of the enthalpy, entropy, free energy, and FEL for protein-solvent systems, need to be introduced.

## 1 Basic concepts of a protein-solvent thermodynamic system

A protein-solvent system is a thermodynamic system composed of the solute (the protein molecules), liquid water, buffer ions, and other substances such as small molecular compounds, cofactors, and ligands. Very complex interactions and energy/heat exchange exist among these substances; therefore, the macroscopic conformational states of a protein and a change in system properties are a result of the average behavior of a very large number of the microscopic constituents. The relationship between these substances and how heat transfer is related to various energy changes within a system are dictated by the laws of thermodynamics [30].

The entropy in a protein-solvent system can be regarded as a measure of how evenly the thermal energy (also called heat, heat energy or kinetic energy in physics) can be distributed over the entire system. Such a concept is central to the second law of thermodynamics, which states that thermal energy always flows spontaneously from regions of higher temperature to regions of lower temperature. Because the tendency to distribute energy as evenly as possible will reduce the state of order of the initial system, entropy is generally treated as an expression of disorder or randomness of the system [6]. When a protein-solvent system is under constant temperature and pressure, the origin of entropy is the heat energy stored in atoms or molecules within the system. This energy makes atoms jostle around and bump onto one another, and this leads to further vibrations, fluctuations, and motions of groups (e.g., amino acid residue side chains), residues, and molecules (e.g., water, protein, and ligands). Conclusively, entropy is a spontaneous process of the transfer and diffusion of thermal energy among various system constituents, which, despite a macroscopic process, originates from microscopic degrees of freedom of atoms and groups, leading ultimately to a as homogeneous as possible energy distribution over the system and, as a result, the random maximization of the system. Entropy cannot be measured directly, but entropy change can be measured quantitatively when it is considered as a change in the degrees of freedom of a system. Entropy change can be further divided into changes in conformational, rotational, and translational entropies of the protein and other solutes, as well as the changes in solvent entropies of the liquid water and ions [5].

Enthalpy is a measure of the total energy of a protein-solvent system. It is the sum of the internal energy (i.e., the energy required to create the system) and the amount of energy required to make room for the system by displacing its environment and establishing its volume and pressure (calculated as the product of the system volume multiplied by the pressure). Enthalpy is a quantifiable state function but the total enthalpy of a protein-solvent system cannot be measured directly. However, the change in enthalpy can be measured when it is expressed as the change in a system's energy. The formations of noncovalent bonds between atoms within a protein and between a protein and the solvent are exothermic processes, and therefore the change in enthalpy is negative. In contrast, breakages of noncovalent bonds are endothermic reactions and therefore lead to a positive enthalpy change of a system. It should be noted that the enthalpy change is a global property of the entire system, which arises from the combined effect of the formations and breakages of various noncovalent bonds (van der Waals contacts, hydrogen bonds, ion pairs, and other polar and nonpolar interactions) between atoms from both the solute and solvent [6].

The thermodynamic free energy is the energy in a physical system that can be converted to do work. For a protein-solvent system, the free energy is generally called the Gibbs free energy [31], which can be regarded as an approximation of the chemical potential, a thermodynamic quantity used to measure the change in the energy of a system. Gibbs free energy is a state function of the system and only when the change in free energy is negative can the state transition process occur spontaneously. The free energy change ( $\Delta G$ ) is expressed as

$$\Delta G = \Delta H - T\Delta S, \quad (1)$$

where  $\Delta H$  and  $\Delta S$  are the change in enthalpy and entropy of the system, respectively, and  $T$  is the temperature in Kelvin. An important principle of thermodynamics is, when a system reaches equilibrium at constant pressure and temperature,  $\Delta G$  is negative and the Gibbs free energy of the system is minimized (the free energy minimum). The corresponding system state at the free energy minimum is called the free energy minimum state. Specifically, no spontaneous process will occur if  $\Delta G$  is positive, unless the system constituents and system conditions are changed, or additional work is done.

For a protein-solvent system, the Gibbs free energy is described further as the FEL [32,33], which is a very useful and vivid way to characterize the states of a protein-solvent system with their corresponding free energy values. Dill [34] defined the FEL of a protein-solvent system as

$$F(x) = F(x_1, x_2, \dots, x_n), \quad (2)$$

where  $F(x)$  is the free energy at a state " $x$ ", which is further characterized by variables  $x_1, x_2, \dots, x_n$  that represent a variety of system states. A system state is the integral state that

results from complex interactions between the solute and solvent, including not only the states of the protein and other small molecules (such as ligand, cofactor, and compound), but also those of water and ions. In protein folding and protein-ligand binding studies, researchers generally focus on the conformational state of the protein, and for simplicity,  $x_1, x_2, \dots, x_n$  are regarded as variables that specify only the microscopic states of the protein [34]. These include all the dihedral angles of the protein chain, the eigenvector projections derived from essential dynamics analysis of an MD trajectory [6,35,36], the number of native contacts, end-to-end distance of the peptide chain, and an order parameter that describes the similarity of the protein structure to the native or other states [37], or any degree of freedom of the protein [34].

## 2 The protein folding problem

The protein folding problem can be divided into two main problems [38,39]: the kinetic question of why the protein can fold so fast, known as Levinthal's paradox [40]; and the thermodynamic question of how the native structural states result from inter-atomic forces acting on an amino acid residue sequence, also called the folding code. The first question was originally posed by Levinthal [40] in 1968 because he noted that although proteins have enormous conformational spaces, they can search, converge and fold quickly to their native states within a microsecond to second timescale. To address the kinetic question, the forces that drive and guide the protein folding process have to be ascertained and elucidated, thus invoking the thermodynamic question, the folding code and its action mechanism. Clearly, these two questions are closely related and both need to be solved to obtain a complete understanding of the protein folding mechanism [6,41].

### 2.1 Kinetics of protein folding

Zwanzig established a mathematical model and solved the kinetic question of protein folding by showing that a small and physically reasonable energy bias against locally unfavorable configurations was enough to reduce Levinthal's time to a biologically significant level [42]. In other words, only when a reasonable free energy gradient exists can the protein fold rapidly into its native state.

Under denaturation conditions, namely, the addition of strong acid/base, a concentrated inorganic salt, or an organic solute (e.g., alcohol, chloroform or urea) to the protein-solvent system, the FEL of the system becomes relatively flat, and the protein loses the tertiary and secondary structures that were present in the native state [43]. If the denatured condition remains unchanged, the unfolded protein molecules wander randomly around the flat free energy surface, resulting in the extensive distribution of a huge

number of conformational states over this endless surface. Therefore, the protein molecules are unable to find out an outlet to become folded. If the renaturation condition is restored, the denatured protein molecules fold rapidly back into their native state, as shown by the famous Anfinsen's experiment that was cited by the Nobel Prize Committee [3,44]. In essence, the recovery of the optimal folding condition changed the shape of FEL and then established a free energy gradient. Such a change is substantially larger than the small free energy fluctuations that occur on the FEL surface under the denatured conditions. Small free energy fluctuations produce a large number of small and shallow energy traps/wells within which different random coiled states are transiently trapped, thus making it easy for conformational transition to occur between these states. Only a substantial change in the shape of the FEL, which allows for the establishment of a steep enough slope, can trigger the folding process irreversibly towards the global free energy minimum state, although numerous traps/wells may also exist within the free energy slope.

### 2.1.1 Funnel-like shape of the FEL for protein folding

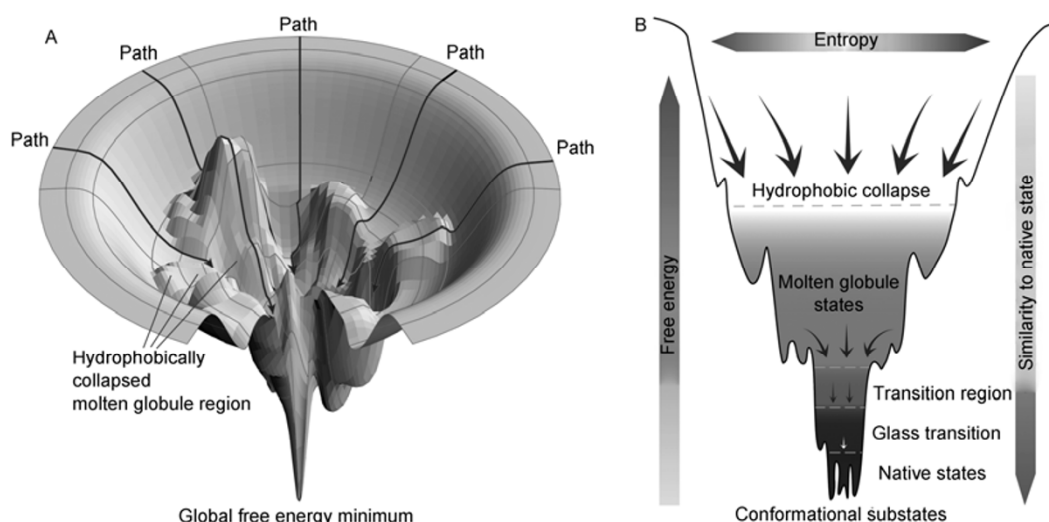
When a steep enough slope in the FEL is established under the proper solvent condition for folding, unfolded peptide chains begin to roll down the free energy slope, just like an ensemble of skiers jumping and skiing down a steep icy slope. This process essentially involves a lowering of the free energy of the protein-solvent system, including the rapid hydrophobic collapse stage [29,45,46] and subsequent slow bottleneck stages [47]. With a decrease in the system

free energy, the protein gradually loses its degree of freedom because of the squeezing effect of water solvent (which is a result of the solvent entropy maximization) and the formation of noncovalent bonds within the protein and between the protein and solvent. The lower the system free energy is, the more conformational entropy the protein loses, and the more ordered conformational states the protein obtains. Thus, the lowering of system free energy accompanied by the reduction in protein conformational freedom makes the shape of the FEL of protein folding look like a funnel [47,48] (Figure 1).

Although the FEL of a protein-solvent system is highly multi-dimensional, it is pictured generally as a three-dimensional surface because it is difficult to draw a multiple-dimensional space. In the canonical depiction of the folding funnel [47,49] (Figure 1), the depth of the funnel (i.e., the difference in free energy between the initial and final states) represents the energetic stability of the native versus the denatured states, and the funnel width represents the conformational space/entropy of the protein. The narrow bottom of the funneled landscape is the global free energy minimum region where the natively folded states of the protein reside. The region outside the funnel is described as a relatively flat surface with many small free energy wells distributed over this surface.

### 2.1.2 Folding process and pathways

Although protein folding problems are complex, several simplified models have been developed to explore and describe the folding process. They include models of diffusion



**Figure 1** Funnel-like FEL for protein folding. A, Schematic three-dimensional representation of the FEL with hills, traps/wells and free energy barriers. Five different potential folding paths are indicated by black lines with arrows showing the downhill directions from the surface outside the funnel (where the denatured states of protein molecules are located) towards the global free energy minimum (where the native states/substates are located). The rough region between the two red lines represents the location of the molten globule state resulting from the hydrophobic collapse process. B, Schematic two-dimensional representation of the folding funnel. The width of the funnel represents the conformational entropy, and the depth represents the energetic stabilization/conformational similarity of the denatured state versus the native state. The arrows indicate the free energy downhill directions and different stages in the folding process. The regions of molten globule states, transition states, glass transition states and native states are color-coded from light blue to dark blue with increasing color depth. A and B have been modified from Dill and Chan [47], and Onuchic et al. [61,62] and Wolynes et al. [63], respectively.

collision [50,51], framework [52], nucleation condensation [53,54], zipping-and-assembly [55–57], Jigsaw puzzle [58], stoichiometry [59,60], hydrophobic collapse [46], and folding funnel [47,48]. These models are not independent and mutually exclusive but are inextricably intertwined and are commonly applied to help understand different aspects of the protein folding process (for details, see [5]). The folding funnel model is based on the FEL theory and embodies most of the aspects characterized by the other models listed above. Therefore, the folding funnel model is the model that is most widely accepted in describing the protein folding process.

Under the folding funnel model, protein folding is viewed as going down a funnel-like FEL via multiple parallel pathways from a vast variety of individual unfolded states to the native states located around the region of the global free energy minimum (Figure 1A) [47]. At any stage the protein molecules exist as an ensemble of conformational states that can be trapped transiently in many local free energy minima because of the ruggedness of the landscape. The rapid equilibrium among these folding intermediates can be disrupted to guarantee a “continuous” downhill movement along different routes as the result of the overall trend in lowering the system free energy. In other words, there may be numerous events during the downhill process that lead to the formation of a variety of folding intermediates such as the molten globule state, transition state, glass transition state, and other relatively unstable states or substates. The detailed folding process for a small globular protein is shown in Figure 1B [61–63]. The first stage is the hydrophobic collapse process, during which an ensemble of unfolded polypeptide chains collapses rapidly into a compact conformational ensemble called the molten globule. The molten globule is an important folding intermediate in which the hydrophobic amino acid side chains are squeezed together into the interior of the protein, some transient secondary structures and nonspecific tertiary interactions are formed, and the protein surface area is minimized. Nevertheless, many native contacts, or close residue-residue interactions present in the native state, have yet to form [62]. A recently published paper [29] reviewed the experimental evidence for (i) the existence of collapsed states under certain conditions (e.g., low concentration of a chemical denaturant) and (ii) the rapid collapse that occurs on a submicrosecond timescale. Each unfolded peptide chain collapses along a different pathway, down the very steep and relatively smooth upper part of the funnel, as shown in Figure 1A, and subsequently enters a rugged region filled with hills, valleys and traps. Molten globule states are therefore trapped and accumulated in these traps/wells/valleys located between hills/ridges for a relatively long time. The subsequent processes are relatively slow steps, because the trapped molten globules need to climb an uphill slope to reach a mountain pass before con-

tinuing to the next downhill search [47]. Such transient trapping, uphill, downhill, transient trapping processes can repeat many times before the next outlet is found, thereby slowing exploration of the routes towards the native state. Therefore, these repetitive processes are regarded as a bottleneck for folding. The ensemble of the conformations located at the hilltop are considered as a “transition state” [47] to highlight the rate aspect of the process, but not the specific structures. Furthermore, a non-complementary change between the entropy and enthalpy can still lower the system free energy, ultimately leading the protein molecules to a so-called glass transition state [32,62]. This trapping state resembles the way a liquid becomes a glass when cooled, remaining fixed in one of many structures and being hard to reconfigure to the lowest free energy state [63]. Therefore, the transition process from the glass transition state to the native state is very slow, requiring a sufficient overall slope of the FEL so that the numerous valleys flow in a funnel towards the native state. In addition, when the folding process reaches the glass transition state, the intermediates have only a few paths towards the native structure [61]. It is worth noting that although the folding funnel model was established originally based on studies of small and fast-folding proteins, it can also be used to describe and explain the folding process of large and complex proteins [64].

## 2.2 Thermodynamics of protein folding

The folding-driving force, regarded as folding code, is encoded in the amino acid sequence because different/similar sequences fold into different/similar structures, and only a fraction of all possible sequences (i.e., the limited number of amino acid sequences found in nature) in the sequence space can fold into functionally structural states. The protein folding code has traditionally been seen as a sum of many individual small interactions including hydrogen bonds, ion pairs, van der Waals attractions, and other polar and nonpolar interactions [39,65]. The idea behind this viewpoint is that the primary sequences encode the protein secondary structures, which then encode the tertiary structures. Intuitively, local interactions lead first to the formation of simple local structures, and then interactions or collisions among these simple structures result in the formation of more complex tertiary structures, which is similar to the hierarchical folding process described by the diffusion collision and framework models [50–52]. However, viewing the sum of many different interactions as the driving force may be one-sided because (i) a dominant component to the folding-driving force, the solvent entropy effect, is ignored [6]; (ii) the statistical mechanical model demonstrated that the folding code is distributed both locally and non-locally in the sequence [38]; and (iii) the local interactions are to a large extent a consequence but not a cause of the folding forces [39].

From a global viewpoint, the folding-driving force is a lowering of total Gibbs free energy of the protein-solvent system, which involves contributions from both the entropy (including changes in the solvent entropy and protein conformational entropy) and enthalpy (including formations and breakages of noncovalent bonds in the protein interior and between the protein and the solvent) [5]. Furthermore, because the protein folding process is very complex, changes in entropy and enthalpy dominate different stages of this process by contributing differentially to the lowering of the system free energy [41].

### 2.2.1 *Why is the solvent entropy effect (the hydrophobic force) important?*

There are large numbers of hydrogen bonds and van der Waals contacts and a relatively small number of ion pairs in already folded proteins. This has led to the idea that these interactions play a dominant role in driving protein folding. Hydrogen bonds are important because all possible hydrogen-bonding interactions are generally satisfied in the native structures, and hydrogen bonds between backbone amide and carbonyl groups are key components of protein secondary structures [39]. Similarly, tight packing in natively folded proteins implies that van der Waals interactions are also important. Compared with the hydrogen bonds and van der Waals contacts, the ion pairs/salt bridges may be less important because most proteins have only a limited number of electrostatically charged residues and many of them are exposed on the solvent accessible surface of proteins. We can conclude that these interactions are clearly important in stabilizing the native conformational states because these noncovalent bonds are observed only in the already folded native structures. However, to what extent they could drive protein folding needs to be examined carefully.

At first glance, the secondary structures are likely to be constrained and stabilized by their key components, the backbone hydrogen bonds. However, isolated secondary structures are seldom stable in solution on their own [39] and, the “chameleon” sequences in natural proteins can assume either  $\alpha$ -helical or  $\beta$ -strand conformations, suggesting that the stability of the secondary structures relies on their tertiary context [66,67], namely, the compact packing of proteins, which is a consequence of the solvent entropy effect that causes hydrophobic collapse during the protein folding process.

The solvent entropy effect, also called hydrophobic force or hydrophobic interaction, has now been widely believed to play a crucial role in protein folding. Direct evidence, as described by Dill [39], is that (i) all natural proteins have hydrophobic cores, implying that the hydrophobic residue side chains are driven to be hidden from water; (ii) transferring a hydrophobic side chain from water into oil-like media requires energy of 1–2 kcal mol<sup>-1</sup> [68]; (iii) proteins are denatured readily in nonpolar solvents; and (iv) sequences with jumbled amino acid residues but correct hydrophobic

and polar patterning can fold to their expected native states [69–71].

All above evidence implies that solvent entropy maximization contributes substantially to the folding-driving force, i.e., the lowering of the total Gibbs free energy of the protein-solvent system. For example, the requirement to maximize the solvent entropy will exclude nonpolar amino acids from water, leading to a minimized protein surface area and a maximized exposure of polar and electrostatically charged residues on the surface. The hydrophobic side chain-solvent system will reach equilibrium, the global free energy minimum state, through solvent entropy maximization and formation of a water shell around the hydrophobic side chain group so as to guarantee the maximal entropy of other water molecules outside the shell. Therefore, the energy is required to disrupt the bonds within the water shell when transferring hydrophobic side chain into the oil-like nonpolar solvent, which is an enthalpically unfavorable process. Interestingly, the process of mixing hydrophobic group with nonpolar solvent is both enthalpically and entropically favorable because the interactions between solute and solvent are easy to establish and the mixture is more disordered than when the hydrophobic group and solvent are not mixed. This process will lead finally to an equilibrium that has a lower system free energy than that of the hydrophobic group-water solvent system. Similarly, the protein denaturation process in nonpolar solvents is also driven by the lowering of the total Gibbs free energy of the protein-nonpolar solvent system, in which both the entropy gain and enthalpy reduction make favorable contributions to the lowering of the system free energy, although the final shape of the FEL exhibits a relatively flat surface [41]. This, from an opposite side, shows that liquid water is an idea solvent that allows the formation and stabilization of the functional conformational states of proteins through its entropy maximization and enthalpy effect (e.g., the formation of hydrogen bonds between the exposed electrostatically charged/polar side chains and the water molecules). The observation that jumbled amino acid residue sequence retaining the correct hydrophobic and polar patterning can fold into the expected native states suggests that the hydrophobic collapse caused by the solvent entropy maximization is the most important stage in protein folding. The resulting molten globule states provide a prerequisite for further sculpting into the native state through conformational adjustments.

### 2.2.2 *Distinct contribution of entropic and enthalpic components to lowering the system free energy*

The statistical mechanic model of the “solvent-induced” force created by Ben-Naim [65] demonstrates that a small conformational change in a group (e.g., the side chain of an amino acid residue) can establish a gradient in the solvation free energy of the protein, which in turn will exert a force on the group, forcing it to move along the direction of the force. In fact, the free energy is a systemic concept that ap-

proximates the overall chemical potential of the protein-solvent system but not merely the solvation free energy of the protein. Furthermore, not only can the conformational change in the protein (i.e., the conformational entropy change of the protein) lead to a change in system free energy, but also the solvent entropy change and the formations and breakages of noncovalent bonds between atoms within the protein and from the protein and solvent (i.e., the enthalpy change), or any change of the solvent condition, e.g., the addition of acid, base, denaturant, cofactor, ligand, substrate or other compounds into the solution, can change the system free energy [41]. The steeper the free energy gradient is, the stronger the force exerted on the protein will be, and the higher the probability that the protein will move towards the lower region of the FEL of the protein-solvent system.

The denaturation process is in essence the process of altering the shape of the FEL from the funnel to a relatively flat surface using denaturants or by changing the system temperature. The renaturation process, as discussed above, is the reverse process of establishing a steep enough free energy gradient that recovers the funneled shape of the FEL through restoring the “normal” solvent condition. The establishment of a steep enough and smooth slope on the upper part of the funnel-like FEL triggers a rapid rolling down of the unfolded protein molecules, resulting finally in the collapse of random coiled states into molten globule states. Traditionally, the hydrophobic interaction was thought to drive such a collapse [38,39,45,46]. However, we consider that the term “hydrophobic interaction” may not be appropriate because it is not a proactive interaction driven by mutual attraction between hydrophobic groups and thus may not be responsible for bringing these groups into the interior of the protein. Indeed, aggregation of hydrophobic side chains into the interior of the natively folded proteins is, as mentioned above, the consequence of the solvent entropy maximization, which is dictated by the second law of thermodynamics.

In the protein-solvent system, the water molecules have an absolute advantage in both quantity and mass compared to the protein molecules [41]. The requirement for the solvent entropy maximization will retain as much as possible the highly dynamic hydrogen bonds among water molecules. This will exclude the hydrophobic residue side chains from contacts with water molecules, thus sequestering them into the interior of the collapsed entities by minimizing the solvent accessible surface area of the entities. During the hydrophobic collapse process, energy is required to strip water molecules from the surface of the denatured protein, which is a positive enthalpy change that makes a negative contribution to the lowering of the system free energy; the conformational entropy of the protein is reduced because of the loss of degree of freedom during the collapse, which also makes a negative contribution to the system free energy lowering. Nevertheless, the solvent entropy gain is over-

whelming compared with the enthalpy increase and conformational entropy loss, and therefore makes a substantial contribution to lowering the total Gibbs free energy of the protein-solvent system. Clearly, the solvent entropy effect is a dominant component of the folding code, and the hydrophobic interactions are the consequence of the hydrophobic collapse rather than the cause for driving protein folding; however, the enthalpy reduction that arises from these hydrophobic contacts can make a minor contribution to lowering the system free energy.

The molten globule that results from the hydrophobic collapse is, as mentioned above, an ensemble of relatively stable folding intermediate states, which are compactly packed and contain some transient secondary structures and tertiary contacts. The molten globule is important because it provides a structural environment for forming not only the local interactions, but also the global tertiary interactions that are present in the natively folded states. As described above, the transition from the molten globule to the native states is a slow process that involves two main folding intermediates: the transition state and the glass transition state. This bottleneck process is slow because of the existence of hills, ridges, valleys, traps/wells, but is still driven by the decrease in total Gibbs free energy of the protein-solvent system. The process of uphill climbing is triggered by the nature of the molten globules to increase their entropy. In other words, the local and/or tertiary interactions are not strong enough to constrain the globules to a particular state. Therefore, the trend to increase the conformational entropy may initially cause the weakest noncovalent bonds to break, disrupting certain structural regions and triggering conformation adjustment, in which noncovalent bonds between atoms are broken and new noncovalent bonds between different atoms are formed. These competitive interactions may repeat many times, until bonds strong enough to counteract the effect of conformational entropy increase are formed. We consider that the conformational entropy increase of a protein may drive the uphill climbing process, while the enthalpy reduction drives the downhill conformational search. As more and more stable noncovalent bonds are formed, the negative enthalpy change will overcompensate for the conformational entropy loss of the protein, thus contributing substantially to the lowering of the system free energy. In addition, the water network formed around the protein surface also makes a favorable contribution to lowering the system free energy. Conclusively, the slow transition process from the molten globule to the native states is still driven by an overall trend in lowering the total Gibbs free energy of the protein-solvent system. Here the negative enthalpy change is a dominant component to the folding code, which overcompensates for the loss of protein conformational entropy and solvent entropy, ultimately leading the protein molecules to arrive at the global free energy minimum region.

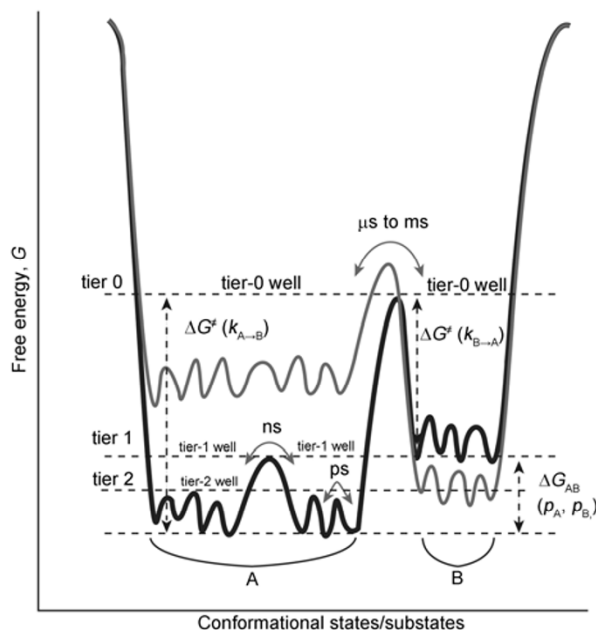
### 3 Physicochemical basis of protein dynamics

When a protein folds into its native structure, we can regard that the protein molecules are at the bottom of the funnel-like FEL of the protein-solvent system. However, the global free energy minimum is not a smooth region and still contains free energy wells/traps and hills/barriers between wells. It is the ruggedness of the FEL bottom that allows the coexistence of different conformational states/substates of protein molecules in an ensemble manner. For example, natively folded protein molecules coexist in equilibrium as ensembles of different conformational states/substates within different free energy wells. The depth and the width of the free energy wells determine the relative populations, probabilities or lifetimes of the conformational states or substates. This is the thermodynamic property of protein dynamics. The height of the free energy barriers that separate two adjacent wells determines the conformational transition rates between two states. This is the kinetic property of the dynamics.

The conformations within individual large free energy wells are not static, but can fluctuate around an average/equilibrium conformational state, resulting in a large ensemble of closely related conformational substates. Because such fluctuations last for a relatively long time (time-scales of microseconds to seconds), they are regarded as equilibrium fluctuations or equilibrium dynamics, which are generally thought to govern the biological functions of proteins [4,6]. On the contrary, the non-equilibrium fluctuations, which occur during the transition process between two conformational states, are thought to have a minimal effect on the overall rates of biological processes [4]. Because conformational transitions need to overcome the free energy barrier and the transition conformational states have transient lifetimes, transition states and the non-equilibrium fluctuations are hard to detect experimentally [5,6].

#### 3.1 Hierarchical dynamics of proteins

As discussed in a number of studies [4,72], protein dynamics have an important fluctuation hierarchy feature, which means that the different structural components in a protein fluctuate with different amplitudes and directionalities on different timescales. The bottom of the funnel-like FEL of a protein-solvent system was shown in Figure 2. The conformational states/substates of a protein reside within individual wells of free energy minimum, with the depth and width of the well determining the amplitude and timescale of the dynamics. The transition states, however, are located on the free energy barriers that separate the wells, with the barrier height determining the directionality of the fluctuations and further, the conformational interconversion rates between two states. The amplitude and timescale of protein dynamics are described below.



**Figure 2** Schematic representation of the funnel-like FEL bottom of a protein-solvent system. The features of the ruggedness, variability and hierarchical organization of the bottom are illustrated. These features dictate not only the fluctuation amplitude and directionality of protein dynamics, but also the hierarchy of fluctuations and their corresponding timescales, thus determining the thermodynamic and kinetic properties of the protein. Conformational states/substates are considered to reside within different free energy wells. For example, the tier-0 states A and B reside within the two largest-and-deepest tier-0 wells, respectively, with the free energy difference  $\Delta G_{AB}$  determining the difference in population distributions of the states A and B (e.g., population  $p_A$  is larger than  $p_B$ ). The difference in free energy barrier between the states A and B ( $\Delta G^\ddagger$ ) determines the conformational interconversion rates (e.g.,  $k_{A \rightarrow B}$  is slower than  $k_{B \rightarrow A}$ ). The tier-0 dynamics involve the whole protein molecule and occur on timescales ranging from microseconds ( $\mu$ s) to milliseconds (ms). Although the tier-0 states A and B coexist in equilibrium with different population distributions, a change in system condition will alter the FEL (from the dark line to the gray line), thus leading to the redistributions of the states A and B. The tier-1 and tier-2 substates reside within tier-1 and tier-2 wells, respectively, and the corresponding fluctuations around these substates involve loops and side chains and occur on faster timescales, e.g., nanoseconds (ns) and picoseconds (ps) for the tier-1 and tier-2 dynamics, respectively. This figure has been modified from Henzler-Wildman and Kern [4] and Ansari et al. [72].

The bottom of the FEL shown in Figure 2 contains two large-and-deep free energy wells, the tier-0 wells. Within the tier-0 wells, there exist relative small or even smaller free energy wells, i.e., two relatively small tier-1 wells exist within the left-side tier-0 well, and four and three smaller tier-2 wells (from left to right) within these two tier-1 wells, respectively. It is such a nested organization of free energy wells that determines the multiple hierarchies of protein dynamics on distinct timescales. For example, the two major conformational states, the tier-0 states, A and B reside within the two largest free energy wells, respectively, and therefore have large populations, long lifetimes or high probabilities ( $p_A$ ,  $p_B$ ). Fluctuations around each tier-0 state (the tier-0 dynamics) are large amplitude motions that in-



volve the whole protein molecule or an entire domain and occur on the timescales ranging from microsecond to million second. Such equilibrium fluctuations are often referred to as large-scale concerted motions or collective motions. The states A and B coexist in equilibrium and the difference in population distributions is determined by their free energy difference  $\Delta G_{AB}$ . For example, the conformational transition from state A to B needs to overcome a higher barrier  $\Delta G^\ddagger (k_{A \rightarrow B})$  than the transition from state B to A ( $\Delta G^\ddagger (k_{B \rightarrow A})$ ). Therefore, the transition rate  $k_{B \rightarrow A}$  will be faster than  $k_{A \rightarrow B}$ , leading to a larger population/higher probability of state A than state B. The fluctuations within the tier-1 wells are fast dynamics (tier-1 dynamics) that occur on timescale of nanoseconds. Such fluctuations involve the loops and turns that connect the secondary structural elements, thus resulting in an ensemble of closely related sub-states within each tier-0 well. Even faster fluctuations (tier-2 dynamics), occur within the smaller tier-2 wells located at the bottom of the tier-1 wells. Such fluctuations mainly involve amino acid side chain rotations on the timescale of picoseconds. These rotations originate from the fastest fluctuations (tier-3 dynamics) such as the bond vibrations that occur on femtosecond timescale.

### 3.2 Relationship between crystal structures and conformational states within the hierarchical free energy wells

A protein structure determined by the X-ray crystallographic method is in essence an average or equilibrium structure under the crystallization condition. Therefore such a structure could be regarded as a conformational state trapped in a tier-0 free energy well. Crystallographic B-factors provide information about the spatial distribution or equilibrium fluctuations around the tier-0 state, indicating that a crystal structure also contains information about the tier-1 and tier-2 states. Therefore, the crystal snapshots and, more importantly, their B-factors are useful for probing the structure-function relationship of a protein.

In the PDB database (<http://www.pdb.org>), there may be several or many crystal structures for the same protein. The conformational differences between these structures are largely minor, because the most pronounced differences are usually in the loop regions rather than in the structural core. This observation implies that (i) the crystallization conditions used by different laboratories were to a large extent similar and therefore trapped similar tier-1 states within the same tier-0 well; and (ii) most of proteins may have only one tier-0 free energy well within which the equilibrium fluctuations that govern the biological functions are found.

A well-known example of a protein having at least two tier-0 states is the HIV-1 gp120 envelope glycoprotein. Significantly large conformational difference are observed between the two states of gp120: the unliganded state and the

CD4-bound state, implying that these two states exist within different tier-0 wells. Most of the gp120 structures determined using X-ray crystallographic methods [73–78], either in the presence or in the absence of the CD4 molecules, are in the CD4-bound state, suggesting that this state is stably trapped in a large and deep tier-0 well. However, the unliganded conformation of gp120 is hard to be crystallized and, therefore, only one unliganded state of SIV gp120 was deposited in the PDB database [79]. This suggests that the unliganded gp120 state is unstable and resides in another tier-0 well that has a higher free energy level than the well within which the CD4-bound state resides. Moreover, the higher conformational flexibility of the unliganded state compared to the CD4-bound state [80,81], suggests that the tier-0 well of the unliganded gp120 may have a more rugged bottom than that of the CD4-bound gp120. In the context of the functional viral spike *in vivo*, however, the unliganded conformation of gp120 can be constrained by interactions with gp41 and other subunits, thus providing advantages for HIV to evade immune surveillance [78]. The conformational stability enhanced by the interactions of gp120 with other molecules indicates that the FEL of the protein-solvent system is not static, but rather, dynamic, with the width and depth of the free energy wells and the height of the barriers being variable under the influence of solvent conditions (Figure 2).

### 3.3 Dynamic nature of the FEL

The free energy wells and barriers shown in Figure 2 are a static manifestation of the dynamic FEL of a protein-solvent system. When the solvent conditions, e.g., the temperature, pressure, pH, ion strength, and the constituents of the system (including the solute and solvent components) are constant, the rugged shape of the landscape bottom will last for a relatively long time period, exhibiting a stable distribution of the hills/barriers and valleys/wells. However, such stability is relative, because the atomic thermal motions and bond vibrations within molecules as well as the Brownian collisions among molecules will inevitably give rise to free energy changes. Such changes will make the rugged profile of the FEL bottom fluctuate around an equilibrium level, similar to the relatively static water ripples over the lake when a breeze is blowing. The existences of stable free energy wells and barriers allow for a stable trap of conformational states, equilibrium fluctuations around the stable states, and stable conformational transition rates between these states, which lead ultimately to a stable distribution of ensembles of conformational states/substates over the rugged bottom of the FEL.

However, any factor capable of perturbing the free energy of the protein-solvent system would likely break such equilibrium and re-establish a new equilibrium, thus leading to the redistribution of the conformational states of the protein molecules [4,41]. Factors that are able to alter the sys-

tem free energy include external and internal ones. External factors include temperature, pressure, pH, ion strength, presence of denaturants, and addition of ligands, cofactors, substrates, compounds or any other molecules into the solution. Internal factors include the amino acid mutations, the effects of protein conformational entropy and solvent entropy, and competitive interactions between protein residues and between residues and solvent. As discussed above, high temperature and denaturants can cause substantial changes in shape of the FEL and thus lead to protein denaturation. These factors can also cause moderate changes in shape of the FEL (Figure 2, gray line), thus leading to the redistribution of protein conformational states. For example, the occasional collision of a ligand with the protein will displace the water molecule network around the surfaces of the two partners, leading to an increase of both the solvent entropy and enthalpy. Further interactions will lead to the formation of noncovalent bonds between the protein and ligand, results in the loss of the system enthalpy and the rotational and translational entropies of the two partners. These non-complementary changes between the entropy and enthalpy will cause the free energy fluctuations of the system, namely, the changes in width and depth of the free energy wells and in height of the free energy barriers.

Non-complementary changes between the entropy and enthalpy originate essentially from the tendency to distribute the heat energy as evenly as possible over the system, that is, the entropy maximization of the protein-solvent system. Atomic thermal energy causes the harmonic oscillations of atoms along the covalent bonds and further, the vibration of the bond connecting two atoms (i.e., the fastest tier-3 dynamics that occur on femtosecond timescale mentioned above). The accumulation of the bond vibrations produces large kinetic energy and the release of such energy causes Brownian motions of water molecules and rotational motions of residue side chains of the protein (i.e., the faster tier-2 dynamics on picosecond timescale). The Brownian collisions of the water molecules can break the hydrogen bonds between two adjacent water molecules and then lead to the formation of new bonds between another two water molecules. The repetition of breakages and formations of hydrogen bonds distributes the water kinetic energy over the system, which ultimately leads to the largest possible number of dynamic hydrogen bonds, that is, the maximization of the solvent entropy. The accumulation of the rotational motions of the protein side chain will break the local noncovalent bonds in regions where the structural constraints are weak (such as the surface-exposed loops and turns). This provides the opportunity to form new noncovalent bonds between protein residues and between residues and water molecules. If the new bonds are not strong enough, the nature to increase conformational entropy can cause these bonds to break, triggering a new round of competitive interactions. This process leads to loop/turn motions on the surface of the protein (i.e., the tier-1 dynamics on nanosec-

ond timescale). Such loop/turn motions are transmitted either through the water molecule network that is formed around the protein surface or through specific structural components (such as the hinge-bending regions) over the entire protein structure, which results in the slow tier-0 dynamics, the collective motions of the protein.

Taken together, the macroscopic dynamics of proteins is a consequence of cascade amplification of the microscopic motions of atoms and atomic groups, for which the entropy originating from atomic thermal energy is most fundamental. Under a constant solvent condition, the trend to increase the system entropy can, to a large extent, be compensated for by an enthalpy increase. Thus, only minor free energy fluctuations occur and this in turn determines the relatively stable distribution of and transition between the different conformational states of a protein. Drastic alterations of the solvent condition, for example the addition of a denaturant into the solution, can introduce large positive entropy that cannot be compensated for by the positive enthalpy. This causes a substantial change in the shape of the FEL and, thus, leads ultimately to protein denaturation. On the contrary, the addition of a ligand or other compounds into the solution usually causes a moderate non-complementary change between the entropy and enthalpy, which leads to the redistribution of the protein conformational states by alterations in the height of free energy barriers and the size of free energy wells.

#### 4 Physicochemical basis underlying the protein-ligand binding

Proteins participate in biological processes and realize their functions mainly through interactions with proteins and peptides, nucleic acids, cofactors, ligands, substrates, small molecule compounds, and other small molecules such as oxygen or metal ions [5,6,82]. Therefore, a detailed understanding of the biological functions of proteins requires an in-depth understanding of the mechanisms underlying protein-ligand recognition and binding and this, in turn, will contribute greatly to drug discovery and design in the field of medicinal chemistry.

##### 4.1 Process of protein-ligand binding and its driving force

The protein-ligand recognition and binding process, like the protein folding process, is driven by the decrease in the total Gibbs free energy of the protein-ligand-solvent system, and this is dictated by a delicately balanced mechanism that involves both the entropy and enthalpy contributions [5,41,83].

Traditionally, the process of protein-ligand binding was described by the “lock-and-key” [84] and “induced-fit” [85] models. The former assumes that both the protein (the lock)

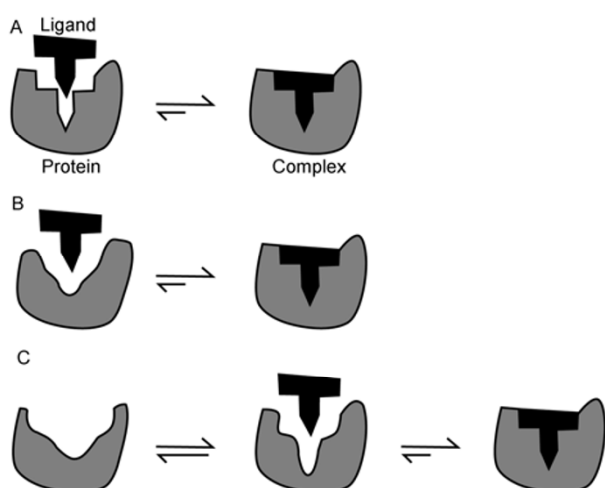
and ligand (the key) are rigid and their binding interfaces are perfectly matched so that the ligand binds to the protein like a key being inserting into a lock (Figure 3A). The latter assumes that the binding interfaces between the protein and ligand are not ideally matched and that the binding-site regions of the protein are flexible. Therefore, the binding of the ligand induces a conformational change in the protein binding site (Figure 3B). These two models have been widely applied to interpret the recognition and interaction mechanisms of enzyme-substrate, target protein-drug, and receptor-ligand binding. The difference in entropy and enthalpy contributions to the binding-driving force between the lock-and-key and induced-fit models is discussed below.

The first step in the protein-ligand binding process is diffusion followed by collisions between the protein and ligand molecules in the solvent. This process is often ignored when binding mechanisms are discussed in literatures. However, it is important because the initial contacts/collisions between two molecules are a prerequisite for further interactions to occur. As discussed above, the release of the water molecule kinetic energy causes Brownian motions of individual water molecules. These motions, on the one hand, can lead to entropy maximization of the water solvent itself or, on the other hand, can cause rotational, translational or diffusion motions of the protein and ligand molecules. It should keep in mind that there is a large quantity of water solvent that surrounds the solute molecules, which can result in strong Brownian bombardments that cause molecule wander and subsequent accidental contacts/collisions between protein and ligand molecules. We emphasize that it is the solvent entropy maximization originating from atomic thermal motions that drives this process. The higher the solute concentrations are, the higher the probability of protein-ligand contact is, which makes it more likely to establish further interactions. This is true for both the lock-and-

key and induced-fit models.

The subsequent steps are different for these two models. For the lock-and-key model, if the initial collisions occur between the complementary interfaces of the protein and ligand, a large number of water molecules will be displaced. Prior to collision, the water molecules formed a well-defined network around the surfaces of the protein and the ligand to suit the requirement for solvent entropy maximization while simultaneously making a favorable contribution to the lowering of the system free energy via the formation of the intermolecular hydrogen bonds (i.e., bonds between the protein and the water and between different water molecules). The initial collision between the protein and ligand will break some of these hydrogen bonds, and this is a process of the positive entropy that originates from the molecular kinetic energy compensating for the positive enthalpy (or energy) that is stored within the water network. This process triggers the overall water displacement and ultimately, leads to the maximization of the solvent entropy upon ligand binding. The ideally complementary interfaces will allow for the displacement of a large number of water molecules, thereby producing a large amount of solvent entropy (the solvent entropy gain) that overcompensates for not only the positive enthalpy change when breaking the hydrogen bonds of the water network, but also the negative entropy change caused by the loss of rotational and translational entropies of the protein and ligand. We conclude that the binding process described by the lock-and-key model is mainly an entropy-driven process, in which the solvent entropy gain makes a substantial contribution to the lowering of the system free energy, just as the hydrophobic collapse does during the protein folding process. When the water molecules are displaced, noncovalent bonds will be formed between the protein and ligand via a negative enthalpy change, which further contributes to the lowering of the system free energy. This process is similar to the slow bottleneck process during protein folding, in which the competitive interactions sculpt the molten globule into the native states. Because the perfect surface complementarity between binding partners is a prerequisite for achieving the large solvent entropy gain, the lock-and-key model can be used to explain the specificity of ligand binding.

For the induce-fit model, the tentative collisions between partners may repeat many times until an appropriate match between the interacting sites is found [86]. Here the “appropriate match” means that, although the contact interfaces are not perfectly complementary, they are well enough matched to provide the initial complex with enough strength and longevity to allow further interactions, which can induce conformational changes in the binding partners. This process will also exclude water molecules from the contact interfaces and thus contribute to the lowering of the system free energy, although this contribution is not as substantial as it is the lock-and-key binding because of the imperfectly complementary interfaces in the induced-fit model. How-



**Figure 3** Schematic representations of models that describe the protein-ligand binding mechanisms. A, Lock-and-key model. B, Induced-fit model. C, Conformational selection model. This figure has been modified from Tobi and Bahar [88].

ever, the high conformational flexibility of the binding-site regions allows for the subsequent conformational adjustments to suit the needs of the incoming ligand and, ultimately, to establish full contact between the interacting partners. The negative enthalpy change originating from the formation of the noncovalent bonds can overcompensate for not only loss of the protein conformational entropy, but also the loss of the rotational and translational entropies of the two binding partners, thereby making a substantial contribution to the lowering of the system free energy. Therefore, we conclude that the binding process described by the induced-fit model is mainly an enthalpy-driven process. Because the large number of non-bonded interactions resulting from the induced fit can stabilize the ligand within the binding site for a long time, we consider that the induced-fit model can be used to explain the ligand binding affinity.

#### 4.2 Conformational selection mechanism of protein-ligand binding

The lock-and-key and induced-fit models can be used to interpret the binding of ligands to rigid and flexible receptor proteins, respectively. These two idealized models focus mainly on the binding process at a single molecule level but not at the population level. As discussed above, protein molecules coexist as ensembles of different conformational states/substates around the rugged FEL bottom of the system. Furthermore, these states/substates can interconvert, with the transition rate determining the probability distribution and population size of these states. Based on the thermodynamic and kinetic properties of proteins, a more “real” model, the conformational selection model [87–89] was proposed to explain the protein-ligand binding mechanism at the molecular population level.

In this model, the ligands can bind selectively to the ensemble of conformational states/substates that has a complementary surface best suiting the ligands, and this is followed by conformational adjustments of the protein, thus shifting the equilibrium towards the complexed/bound state (Figure 3C). These two consecutive steps are similar to the two processes described by the lock-and-key and induced-fit models, respectively, and thus are driven by both solvent entropy gain and system enthalpy loss. In the conformational selection model, unliganded and complexed states (or approximately complexed states) coexist in equilibrium at the bottom of the FEL, while selective binding to the complexed states disrupts such equilibrium by altering the height of the free energy barrier (or the conformational transition rates), thus leading to a redistribution of these states with altered population size. In the binding process described by the conformational selection model, selective binding can not only efficiently lower the entropy penalty (e.g., the inevitable loss of the protein conformational entropy in the induced-fit model) but can also gain as much as possible the solvent entropy (e.g., as in the lock-and-key

model). In nature, most proteins are flexible, especially in their ligand-binding regions. This allows protein molecules to exist as an ensemble of closely related conformational substates, and also allows conformational adjustments to occur to establish full contact between the two binding partners.

Thus, the conformational selection model, which describes the binding process as a selective binding followed by an induced-fit, is driven by both the solvent entropy gain and system enthalpy loss. Because this model (i) includes the binding processes described by the lock-and-key and induced-fit models, (ii) is based on the FEL theory and hence is capable of describing the binding process at the molecular population level, and (iii) can interpret both the ligand-binding specificity and affinity, we consider that it is a more comprehensive and realistic model for explaining and describing protein-ligand binding.

#### 4.3 “Misfolding disease” is in fact “protein-protein binding” disease

Alzheimer’s disease and prion diseases such as “Mad Cow”, scrapie, and Creutzfeldt-Jakob diseases [90] are called “misfolding diseases” because the symptoms are believed to be caused by the misfolding of relevant proteins and the aggregation of these misfolded proteins in the brain and other tissues.

Experimental evidence has shown that certain structural regions of the amyloid beta (A $\beta$ ) protein and the prion protein can convert from their “native”  $\alpha$ -helical conformation to a “natively unfolded”, prone-to-plaque/scrapie-formation  $\beta$ -stranded conformation [91,92]. This implies that the bottoms of the FELs of these two proteins are rugged, allowing for the co-existence of their native and natively unfolded states [93]. The natively unfolded states may reside in a relatively smaller free energy well with a higher free energy level than the well within which the native states reside. Therefore, the natively unfolded states have a smaller population with shorter lifetimes than the native states. Nevertheless, the flexible surfaces of the natively unfolded conformers provide an opportunity for recognition and binding between these conformers, thereby forming seed for fibril growth. Once the seed is formed, the fibril will grow rapidly as the result of a large free energy decrease that arises from the solvent entropy gain. The formation of noncovalent bonds between the binding interfaces will contribute further to the lowering of the free energy, ultimately deepening and widening the free energy wells of the natively unfolded states, and shifting the equilibrium towards aggregation complex, as seen in the plaque observed in the tissues of patients. Therefore, the “misfolded” conformations that are finally observed in the plaque are more a consequence than a cause of the protein aggregation.

We speculate that it may be accumulative mutations that loosen the structural constraints in certain regions of the A $\beta$

and prion proteins, thereby altering the shape of the FEL through the increased conformational entropy, which in turn results in the coexistence of at least two conformational states (i.e., two tier-0 states) of the proteins. Aggregation is merely a manifestation of the side-effect of protein binding. In other words, the higher-energy, less populated conformationally altered monomeric states can recognize and bind selectively to one another to form the lower-energy, highly polymorphic aggregate species [94]. Although the misfolding diseases are caused by the accumulation of misfolded proteins which could not be cleared effectively by the protein quality control system in cells, we emphasize that it is the inter-molecular binding/interaction between proteins that stabilizes the misfolded states and protect them from the protein quality control system. Interestingly, because of the dynamic variability of the FEL, manipulating the solvent conditions or factors that can affect the shape of the FEL may help the aggregated proteins “jump out of” the free energy wells, thus shedding light on possible ways to treat the “misfolding disease” [41].

## 5 Protein folding funnel and binding funnels: concluding remarks

Protein folding and protein-ligand binding are similar processes, both being driven by the decrease in total Gibbs free energy of the systems. The only difference between them is the chain connectivity of the system components, giving rise to two different terms: intramolecular and intermolecular recognition and binding [5,95,96]. However, whether the recognition and binding occur within a molecule or between different molecules, the decrease in system free energy, which originates from the non-complementary change between the enthalpy and the entropy, drives the formation of stable, natively folded conformational states of a protein or a protein-ligand complex [97].

For protein folding, the lowering of the system free energy coupled with the gradual reduction in conformational degree of freedom of folding intermediates, determines that the FEL for protein folding must be funnel-like. Although protein-ligand associations occur around the rugged bottom of the FEL, the exclusion of water from the binding interfaces and the formation of noncovalent bonds between the two partners can still lower the system free energy. In conjunction with the loss of the rotational and translational degrees of freedom of the binding partners as well as the loss of the conformational entropy of the protein, these processes could merge, downwards expand, and further widen the free energy wells within which the protein-ligand binding process takes place, thereby making them look like a funnel, which we called the binding funnel [97].

Based on the FEL theory, Nussinov et al. proposed a building block model to demonstrate the similarity between protein folding and protein-ligand binding [95,98]. In this

model, the protein or the ligand was divided into a set of building blocks that were located within different microfunnel-like free energy wells. Then, the processes of protein folding or protein-ligand binding were considered as the recognition and association among these building blocks that are driven by fusing microfunnels into a higher dimensional funnel, regardless of the chain connectivity. Therefore, the essence of the building block model is a series of microfunnel fusion events, which lower the total Gibbs free energy of the system and lead to the global free energy minimum states of the protein or the protein-ligand complex.

The free energy downhill processes for protein folding and protein-ligand binding are also similar. In the protein folding process, the first stage is the hydrophobic collapse driven by the solvent entropy maximization, resulting in the molten globule intermediate in which some of the native secondary structural elements and tertiary contacts may have been formed, while many native contacts or close residue-residue interactions present in the native state have yet to form. We consider that the molten globule states are important because they provide a structural environment for further conformational adjustments through conformational entropy increase and competitive interactions between protein residues and between residues and water molecules. In the protein-ligand binding process, the first step is also driven by the solvent entropy effect. This includes two consecutive sub-steps: (i) the trend to increase the solvent entropy causes solute molecules to wander and promotes the subsequent collisions between them, and (ii) the requirement to maximize the entropy displaces the water molecule network around the collision interfaces of the two partners. Therefore, these two sub-steps are also driven by the solvent entropy maximization, ultimately resulting in an initial complex within which some noncovalent bonds have been established, although the two partners are still loosely associated. The initial complex is analogous to the molten globule because it provides the structural environment for further sculpting into the final compactly packed/tightly bound states. The second step is driven mainly by the negative enthalpy change through competitive interactions triggered by conformational entropy increase. Therefore, whether in the protein folding or in the protein-ligand binding process, both the entropy-driven first step and the enthalpy-driven second step contribute to the lowering of the system free energy, which results in the folding and binding FEL being funnel-like.

The non-complementary change between entropy and enthalpy will bring about free energy fluctuations that lead to the ruggedness either in the funnel wall or around the bottom of the funnel. For protein folding, the overall trend in free energy reduction overcomes the negative effect of the ruggedness in the funnel wall, making it possible for the unfolded protein to roll down towards the global free energy minimum region. When protein molecules arrive at the bot-

tom of the funneled FEL, the ruggedness allows for the co-existence in equilibrium of ensembles of different conformational states/substates with relatively stable population distributions (the thermodynamic property of the protein dynamics) and conversion rates (the kinetic property). However, alternations of the solvent conditions (e.g., the addition of the ligand), or internal changes in the protein (e.g., the amino acid mutation) will perturb the system free energy and disrupt this equilibrium, finally resulting in re-distributions of the conformational states/substates.

Although the fluctuations/perturbations of the system free energy come from the non-complementary change between the entropy and enthalpy, the trend to increase the entropy is most primary and critical in triggering the free energy change of the system [99]. This is because (i) during the protein denaturation process, the denaturant molecules/the increasing temperature initially attack/melts the weakest constrained regions of the protein that have large conformational entropy [100], e.g., the surface-exposed loops and turns; (ii) during the protein folding and protein-ligand binding processes, the requirement for solvent entropy maximization establishes a free energy gradient that is steep enough to trigger the hydrophobic collapse and promote the formation of the initial complex; furthermore, the nature to increase the conformational entropy of the protein triggers the conformational adjustments via competitive interactions; and (iii) for the natively folded protein molecules, the nature to increase the conformational entropy makes the less well constrained regions fluctuate around the protein surface, resulting in a large ensemble of closely related conformational substates. These fluctuations, together with the fluctuations of the water network (which originate from the requirement to maximize the solvent entropy) around the protein surface, can transmit over the entire protein molecule, leading to the large concerted motions that are most relevant to function [6,36,80,81,101–103]. Interestingly, using the relative entropy as a minimization object function has been shown to be an effective approach for protein design, indicating the importance of entropy in both folding and inverse folding of proteins [104].

In summary, we consider that the tendency to maximize the entropy of the protein-solvent system that originates from the atomic thermal energy is the most fundamental driving factor for protein folding, binding and dynamics. However, the enthalpy reduction, an opposing factor that tends to make the system become ordered, can compensate for the effect of entropy, which allows the system to reach equilibrium at the global or local free energy minimum.

*This work was supported by the National Natural Science Foundation of China (31370715, 31160181, 31360277, 30860011), the National Basic Research Program of China (2013CB127500), the Program of Innovation Group of Yunnan Province (2011CI123), and Foundation for Key Teacher in Yunnan University (XT412003).*

- 1 Wright PE, Dyson HJ. Intrinsically unstructured proteins: reassessing the protein structure-function paradigm. *J Mol Biol*, 1999, 293: 321–331
- 2 Anfinsen CB. The formation and stabilization of protein structure. *Biochem J*, 1972, 128: 737–749
- 3 Anfinsen CB. Principles that govern the folding of protein chains. *Science*, 1973, 181: 223–230
- 4 Henzler-Wildman KA, Kern D. Dynamic personalities of proteins. *Nature*, 2007, 450: 964–972
- 5 Liu SQ, Xie YH, Ji XL, Tao Y, Tan DY, Zhang KQ, Fu YX. Protein folding, binding and energy landscape: a synthesis. In: Kaumaya PTP, ed. *Protein Engineering*. Rijeka: Intech, 2012. 207–252
- 6 Yang LQ, Sang P, Xie YH, Tao Y, Fu YX, Zhang KQ, Liu SQ. Protein dynamics and motions in relation to their functions: Several case studies and the underlying mechanisms. *J Biomol Struct Dyn*, 2014, 32: 372–393
- 7 Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nat Struct Biol*, 2002, 9: 646–652
- 8 Hansson T, Oostenbrink C, van Gunsteren W. Molecular dynamics simulations. *Curr Opin Struct Biol*, 2002, 12: 190–196
- 9 Kannan S, Zacharias M. Simulated annealing coupled replica exchange molecular dynamics—an efficient conformational sampling method. *J Struct Biol*, 2009, 166: 288–294
- 10 Haliloglu T, Bahar I, Erman B. Gaussian dynamics of folded proteins. *Phys Rev Lett*, 1997, 79: 3090–3093
- 11 Scheraga HA, Khalili M, Liwo A. Protein-folding dynamics: overview of molecular simulation techniques. *Annu Rev Phys Chem*, 2007, 58: 57–83
- 12 Laio A, Parrinello M. Escaping free-energy minima. *Proc Natl Acad Sci USA*, 2002, 99: 12562–12566
- 13 Granata D, Camilloni C, Vendruscolo M, Laio A. Characterization of the free-energy landscapes of proteins by nmr-guided metadynamics. *Proc Natl Acad Sci USA*, 2013, 110: 6817–6822
- 14 Laio A, Gervasio FL. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep Prog Phys*, 2008, 71: 126601–126623
- 15 Kastner J. Umbrella sampling. *Wiley Interdiscip Rev Comput Mol Sci*, 2011, 1: 932–942
- 16 Torrie GM, Valleau JP. Non-physical sampling distributions in monte-carlo free-energy estimation: umbrella sampling. *J Comput Phys*, 1977, 23: 187–199
- 17 Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. Multidimensional free-energy calculations using the weighted histogram analysis method. *J Comput Chem*, 1995, 16: 1339–1350
- 18 Roux B. The calculation of the potential of mean force using computer simulations. *Comput Phys Comm*, 1995, 91: 275–282
- 19 Adcock SA, McCammon JA. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem Rev*, 2006, 106: 1589–1615
- 20 Paci E, Karplus M. Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. *J Mol Biol*, 1999, 288: 441–459
- 21 Schlitter J, Engels M, Kruger P, Jacoby E, Wollmer A. Targeted molecular-dynamics simulation of conformational change—application to the T-R transition in insulin. *Mol Simul*, 1993, 10: 291–308
- 22 Jarzynski C. Nonequilibrium equality for free energy differences. *Phys Rev Lett*, 1997, 78: 2690–2693
- 23 Stone JE, Phillips JC, Freddolino PL, Hardy DJ, Trabuco LG, Schulten K. Accelerating molecular modeling applications with graphics processors. *J Comput Chem*, 2007, 28: 2618–2640
- 24 Brown WM, Kohlmeyer A, Plimpton SJ, Tharrington AN. Implementing molecular dynamics on hybrid high performance computers: particle-particle particle-mesh. *Comput Phys Comm*, 2012, 183: 449–459
- 25 Harvey M, Giupponi G, De Fabritiis G. Acemd: accelerated molecular dynamics simulations in the microseconds timescale. *J Chem Theory Comput*, 2009, 5: 1632–1639
- 26 Liu SQ, Liang LM, Tao Y, Yang LQ, Ji XL, Yang JK, Fu YX, Zhang KQ. Structural and dynamic basis of serine proteases from nema-

- trophagous fungi for cuticle degradation. In: Stoytcheva M, ed. *Pesticides in the Modern World—Pests Control and Pesticides Exposure and Toxicity Assessment*. Rijeka: Intech, 2011. 333–376
- 27 Rapaport DC. *The Art of Molecular Dynamics Simulation*. Cambridge: Cambridge University Press, 1998
  - 28 Sagui C, Daren TA. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annu Rev Biophys Biomol Struct*, 1999, 28: 155–179
  - 29 Haran G. How, when and why proteins collapse: the relation to folding. *Curr Opin Struct Biol*, 2012, 22: 14–20
  - 30 Haynie DT. *Biological Thermodynamics*. Cambridge: Cambridge University Press, 2001
  - 31 Gibbs JW. A method of geometrical representation of the thermodynamic properties of substances by means of surfaces. *Trans Conn Acad Arts Sci*, 1873, 2: 382–404
  - 32 Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Protein Struct Funct Genet*, 1995, 21: 167–195
  - 33 Wales DJ. *Energy Landscapes*. Cambridge: Cambridge University Press, 2003
  - 34 Dill KA. Polymer principles and protein folding. *Protein Sci*, 1999, 8: 1166–1180
  - 35 Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. *Protein Struct Funct Genet*, 1993, 17: 412–425
  - 36 Tao Y, Rao ZH, Liu SQ. Insight derived from molecular dynamics simulation into substrate-induced changes in protein motions of proteinase K. *J Biomol Struct Dyn*, 2010, 28: 143–157
  - 37 Kapon R, Nevo R, Reich Z. Protein energy landscape roughness. *Biochem Soc Trans*, 2008, 36: 1404–1408
  - 38 Dill KA, Ozkan SB, Weikl TR, Chodera JD, Voelz VA. The protein folding problem: when will it be solved? *Curr Opin Struct Biol*, 2007, 17: 342–346
  - 39 Dill KA, Ozkan SB, Shell MS, Weikl TR. The protein folding problem. *Annu Rev Biophys Biomol Struct*, 2008, 37: 289–316
  - 40 Levinthal C. Are there pathways for protein folding? *J Chim Phys*, 1968, 65: 44–45
  - 41 Yang LQ, Ji XL, Liu SQ. The free energy landscape of protein folding and dynamics: a global view. *J Biomol Struct Dyn*, 2013, 31: 982–992
  - 42 Zwanzig R, Szabo A, Bagchi B. Levinthal's paradox. *Proc Natl Acad Sci USA*, 1992, 89: 20–22
  - 43 Li W, Qin M, Tie Z, Wang W. Effects of solvents on the intrinsic propensity of peptide backbone conformations. *Phys Rev E*, 2011, 84: 041933
  - 44 Anfinsen CB, Haber E, Sela M, White FHJ. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci USA*, 1961, 47: 1309–1314
  - 45 Agashe VR, Shastry MCR, Udgaonkar JB. Initial hydrophobic collapse in the folding of barstar. *Nature*, 1995, 377: 754–757
  - 46 Dill KA. Theory for the folding and stability of globular proteins. *Biochemistry*, 1985, 24: 1501–1509
  - 47 Dill KA, Chan HS. From levinthal to pathways to funnels. *Nat Struct Biol*, 1997, 4: 10–19
  - 48 Leopold PE, Montal M, Onuchic JN. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc Natl Acad Sci USA*, 1992, 89: 8721–8725
  - 49 Dobson CM. The nature and significance of protein folding. In: Pain RH, ed. *Mechanisms of Protein Folding*. 2nd ed. Oxford: Oxford University Press, 2000. 1–33
  - 50 Karplus M, Weaver DL. Protein-folding dynamics. *Nature*, 1976, 260: 404–406
  - 51 Karplus M, Weaver DL. Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci*, 1994, 3: 650–668
  - 52 Kim PS, Baldwin RL. Specific intermediates in the folding reactions of small proteins and the mechanism of folding. *Annu Rev Biochem*, 1982, 51: 459–489
  - 53 Nolting B, Agard DA. How general is the nucleation-condensation mechanism? *Proteins Struct Funct Bioinf*, 2008, 73: 754–764
  - 54 Yan Z, Wang J, Zhang Y, Qin M, Wang W. Nucleation process in the folding of a domain-swapped dimer. *Phys Rev E*, 2010, 81: 021910
  - 55 Dill KA, Fiebig KM, Chan HS. Cooperativity in protein-folding kinetics. *Proc Natl Acad Sci USA*, 1993, 90: 1942–1946
  - 56 Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. Principles of protein folding—a perspective from simple exact models. *Protein Sci*, 1995, 4: 561–602
  - 57 Sun L, Wang J, Wang W. Dissection of the Zipping-and-Assembly mechanism for folding of model proteins. *Chin Phys Lett*, 2010, 27: 038702
  - 58 Harrison SC, Durbin R. Is there a single pathway for the folding of a polypeptide chain? *Proc Natl Acad Sci USA*, 1985, 82: 4028–4030
  - 59 Mittal A, Jayaram B, Shenoy SR, Bawa TS. A stoichiometry driven universal spatial organization of backbones of folded proteins: Are there chagaff's rules for protein folding? *J Biomol Struct Dyn*, 2010, 28: 133–142
  - 60 Ji XL, Liu SQ. Is stoichiometry-driven protein folding getting out of thermodynamic control? *J Biomol Struct Dyn*, 2011, 28: 621–623
  - 61 Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. *Ann Rev Phys Chem*, 1997, 48: 545–600
  - 62 Onuchic JN, Wolynes PG, Luthey-Schulten Z, Socci ND. Towards an outline of the topography of a realistic protein folding funnel. *Proc Natl Acad Sci USA*, 1995, 92: 3626–3630
  - 63 Wolynes PG, Onuchic JN, Thirumalai D. Navigating the folding routes. *Science*, 1995, 267: 1619–1620
  - 64 Li W, Terakawa T, Wang W, Takada S. Energy landscape and multi-route folding of topologically complex proteins adenylate kinase and 2ouf-knot. *Proc Natl Acad Sci USA*, 2012, 109: 17789–17794
  - 65 Ben-Naim A. Levinthal's question revisited, and answered. *J Biomol Struct Dyn*, 2012, 30: 113–124
  - 66 Mezei M. Chameleon sequences in the pdb. *Protein Eng*, 1998, 11: 411–414
  - 67 Minor DL, Kim PS. Context-dependent secondary structure formation of a designed protein sequence. *Nature*, 1996, 380: 730–734
  - 68 Wolfenden R. Experimental measures of amino acid hydrophobicity and the topology of transmembrane and globular proteins. *J Gen Physiol*, 2007, 129: 357–362
  - 69 Cordes MHJ, Davidson AR, Sauer RT. Sequence space, folding and protein design. *Curr Opin Struct Biol*, 1996, 6: 3–10
  - 70 Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH. Protein design by binary patterning of polar and nonpolar amino acids. *Science*, 1993, 262: 1680–1685
  - 71 Hecht MH, Das A, Go A, Bradley LH, Wei Y. *De novo* proteins from designed combinatorial libraries. *Protein Sci*, 2004, 13: 1711–1723
  - 72 Ansari A, Berendzen J, Bowne SF, Frauenfelder H, Iben IET, Sauke TB, Shyamsunder E, Young RD. Protein states and protein quakes. *Proc Natl Acad Sci USA*, 1985, 82: 5000–5004
  - 73 Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structure of an HIV-1 gp120 envelope glycoprotein in complex with the cd4 receptor and a neutralizing human antibody. *Nature*, 1998, 393: 648–659
  - 74 Kwong PD, Wyatt R, Majeed S, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. *Structure*, 2000, 8: 1329–1339
  - 75 Huang CC, Tang M, Zhang MY, Majeed S, Montabana E, Stanfield RL, Dimitrov DS, Korber B, Sodroski J, Wilson IA, Wyatt R, Kwong PD. Structure of a V3-containing HIV-1 gp120 core. *Science*, 2005, 310: 1025–1028
  - 76 Huang CC, Lam SN, Acharya P, Tang M, Xiang SH, Hussan SS, Stanfield RL, Robinson J, Sodroski J, Wilson IA, Wyatt R, Bewley CA, Kwong PD. Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. *Science*, 2007, 317: 1930–1934
  - 77 Pancera M, Majeed S, Ban YA, Chen L, Huang CC, Kong L, Kwon YD, Stuckey J, Zhou T, Robinson JE, Schief WR, Sodroski J, Wyatt R, Kwong PD. Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conforma-

- tional mobility. *Proc Natl Acad Sci USA*, 2010, 107: 1166–1171
- 78 Kwon YD, Finzi A, Wu X, Dogo-Isonagie C, Lee LK, Moore LR, Schmidt SD, Stuckey J, Yang Y, Zhou T, Zhu J, Vicic DA, Debnath AK, Shapiro L, Bewley CA, Mascola JR, Sodroski JG, Kwong PD. Unliganded HIV-1 gp120 core structures assume the CD4-bound conformation with regulation by quaternary interactions and variable loops. *Proc Natl Acad Sci USA*, 2012, 109: 5663–5668
  - 79 Chen B, Vogan EM, Gong H, Skehel JJ, Wiley DC, Harrison SC. Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature*, 2005, 433: 834–841
  - 80 Liu SQ, Fu YX, Liu CQ. Molecular motions and conformational transition between different conformational states of HIV-1 gp120 envelope glycoprotein. *Chin Sci Bull*, 2007, 52: 3074–3088
  - 81 Liu SQ, Liu SX, Fu YX. Molecular motions of human HIV-1 gp120 envelope glycoproteins. *J Mol Model*, 2008, 14: 857–870
  - 82 Perozzo R, Folkers G, Scapozza L. Thermodynamics of protein-ligand interactions: history, presence, and future aspects. *J Recept Signal Transduct Res*, 2004, 24: 1–52
  - 83 Ji XL, Liu SQ. Thinking into mechanism of protein folding and molecular binding. *J Biomol Struct Dyn*, 2011, 28: 995–996
  - 84 Fischer E. Einfluss der configuration auf die wirkung der enzyme. *Ber Dtsch Chem Ges*, 1894, 27: 2984–2993
  - 85 Koshland DEJ. Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci USA*, 1958, 44: 98–104
  - 86 Bosshard HR. Molecular recognition by induced fit: How fit is the concept? *News Physiol Sci*, 2001, 16: 171–173
  - 87 Foote J, Milstein C. Conformational isomerism and the diversity of antibodies. *Proc Natl Acad Sci USA*, 1994, 91: 10370–10374
  - 88 Tobi D, Bahar I. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc Natl Acad Sci USA*, 2005, 102: 18908–18913
  - 89 Changeux JP, Edelstein S. Conformational selection or induced fit? 50 years of debate resolved. *F1000 Biol Rep*, 2011, 3: 19
  - 90 Serban D, Taraboulos A, DeArmond SJ, Prusiner SB. Rapid detection of Creutzfeldt-Jakob disease and scrapie prion proteins. *Neurology*, 1990, 40: 110–117
  - 91 Kosik KS. Alzheimer's disease: a cell biological perspective. *Science*, 1992, 256: 780–783
  - 92 Pan KM, Baldwin M, Nguyen J, Gasset M, Serban A, Groth D, Mehlhorn I, Huang Z, Fletterick RJ, Cohen FE, Prusiner SB. Conversion of  $\alpha$ -helices into  $\beta$ -sheets features in the formation of the scrapie prion proteins. *Proc Natl Acad Sci USA*, 1993, 90: 10926–10966
  - 93 Weinreb PH, Zhen W, Poon AW, Conway KA, Lansbury PTJ. Nacp, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry*, 1996, 35: 13709–13715
  - 94 Csermely P, Palotai R, Nussinov R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci*, 2010, 35: 539–546
  - 95 Ma B, Kumar S, Tsai CJ, Nussinov R. Folding funnels and binding mechanisms. *Protein Eng*, 1999, 12: 713–720
  - 96 Tsai CJ, Xu D, Nussinov R. Protein folding via binding and vice versa. *Fold Des*, 1998, 3: R71–R80
  - 97 Xie YH, Sang P, Tao Y, Liu SQ. Protein folding and binding funnels: a common driving force and a common mechanism. *J Biomol Struct Dyn*, 2013, 31(suppl 1): 100–101
  - 98 Tsai CJ, Kumar S, Ma B, Nussinov R. Folding funnels, binding funnels, and protein function. *Protein Sci*, 1999, 8: 1181–1190
  - 99 Xie YH, Tao Y, Liu SQ. Wonderful roles of the entropy in protein dynamics, binding and folding. *J Biomol Struct Dyn*, 2013, 31(suppl 1): 98–100
  - 100 Xu X, Su J, Chen W, Wang C. Thermal stability and unfolding pathways of Sso7d and its mutant F31A: insight from molecular dynamics simulation. *J Biomol Struct Dyn*, 2011, 28: 717–727
  - 101 Liu SQ, Liu CQ, Fu YX. Molecular motions in HIV-1 gp120 mutants reveal their preferences for different conformations. *J Mol Graphics Modell*, 2007, 26: 306–318
  - 102 Liu SQ, Meng ZH, Fu YX, Zhang KQ. Insights derived from molecular dynamics simulation into the molecular motions of serine protease proteinase K. *J Mol Model*, 2010, 16: 17–28
  - 103 Liu S Q, Meng Z H, Fu Y X, Zhang K Q. The effect of calciums on the molecular motions of proteinase K. *J Mol Model*, 2011, 17: 289–300
  - 104 Liu Y, Wang B, Wang C, Chen W. A new approach for protein design based on the relative entropy. *Sci China Ser G*, 2003, 46: 659–669